

# EXHIBIT B

**From:** [Joshua Stein](#)  
**To:** [Lauter, Judd](#); [Dunning, Angela L.](#); [Kathleen Hartnett](#); [mlemley@lex-lumina.com](mailto:mlemley@lex-lumina.com); [Ghajar, Bobby A.](#); [Ghazarian, Colette A](#)  
**Cc:** [Jesse Panuccio](#); [Poppell, Cole A](#); [Biksa, Liene](#); [Weinstein, Mark](#); [Stameshkin, Liz](#); [Alvarez, Jessica](#); [Holden Benon](#); [Christopher Young](#); [Aaron Cera](#); [Cadio Zirpoli](#); [Joe Saveri](#); [Margaux Poueymirou](#); [Ashleigh Jensen](#); [Rya Fishman](#); [Matthew Butterick](#); [Nada Djordjevic](#); [James Ulwick](#); [Bryan L. Clobes](#); [Mohammed Rathur](#); [Amy Keller](#); [David Straite](#); [Ruby Ponce](#); [Alexander Sweatman](#); [Heaven Haile](#); [Llama BSF](#); [Josh Schiller](#); [David Boies](#); [Maxwell Pritt](#); [z/Meta-Kadrey](#)  
**Subject:** Letter to Meta re Pending Discovery Issues  
**Date:** Wednesday, October 9, 2024 5:51:04 AM  
**Attachments:** [10092024 Letter to Meta re Pending Discovery Issues.pdf](#)

---

Counsel,  
Please see the attached correspondence.  
Best,  
Josh

**Joshua Michelangelo Stein**

Partner

---

**BOIES SCHILLER FLEXNER LLP**

44 Montgomery Street, 41<sup>st</sup> Floor

San Francisco, CA 94104

(t) +1 415 293 6813

(m) +1 617 365 3991

[jstein@bsfllp.com](mailto:jstein@bsfllp.com)

[www.bsfllp.com](http://www.bsfllp.com)



October 9, 2024

**SENT VIA EMAIL**

Maxwell V. Pritt  
Boies Schiller Flexner LLP  
44 Montgomery Street, 41st Floor  
San Francisco, CA 94104  
mpritt@bsfllp.com  
(415) 293-6800

Bobby Ghajar  
bghajar@cooley.com  
Colette Ghazarian  
cghazarian@cooley.com  
Cooley LLP  
133 2nd Street, Suite 400  
Santa Monica, California 90401

Angela L. Dunning  
adunning@cooley.com  
1841 Page Mill Road, Suite 250  
Palo Alto, CA 94304

Mark Weinstein  
mweinstein@cooley.com  
Kathleen Hartnett  
khartnett@cooley.com  
Judd Lauter  
jlauter@cooley.com  
Elizabeth L. Stameshkin  
estameshkin@cooley.com  
Cooley LLP  
3175 Hanover Street  
Palo Alto, CA 94304-1130

**Re: *Kadrey v. Meta*, Case No. 3:23-cv-03417-VC,  
Pending Disputes Over Meta's Discovery Responses and Productions**

Dear Counsel,

We write concerning several unresolved deficiencies that Plaintiffs previously raised, in whole or in part, with Meta concerning Meta's objections and responses to Plaintiffs' discovery requests and Meta's document productions. In the interest of resolving these issues without court intervention, we identify the specific deficiencies below and request that Meta confirm by 5 p.m. pacific on Thursday, October 10, whether it will supplement its responses and productions as requested herein.

For any deficiency Meta declines to correct, we must proceed to raise these issues with Judge Hixson in accordance with his Discovery Standing Order.

**I. Requests for Production**



RFP No. 44:

This RFP requests all documents Meta provided to a “governmental entity, elected officeholder, or appointed official, including state or local regulatory Documents, voluntarily or otherwise, regarding generative AI, Training Data, the Meta Language Models, and all related Communications.” The parties met and conferred on April 24, 2024, but this issue remains unresolved. Plaintiffs request Meta promptly produce all responsive documents or confirm in writing if none exist.

RFP No. 45:

During the parties’ meet and confer on June 21, 2024, Meta agreed to conduct “a reasonable search for documents within Meta’s possession, custody, or control that are responsive.” In an August 22, 2024 letter, Plaintiffs indicated that their document review showed Meta had created a process for identifying potential licensors and reaching out to business development partners, yet Meta has not produced any documents or communications between Meta researchers, business development partners, or content companies regarding potential licensing. Plaintiffs requested that Meta search for and produce those documents.

During the parties’ meet and confer on August 29, 2024, in response to Plaintiffs’ request for additional documents regarding licensing and communications with textbook publishers and literary works publishers, Meta argued that the specific language of RFP No. 45 clearly refers only to “some kind of mechanism or tool that would keep track of or provide for licensing or accreditation or attribution,” and does not broadly encompass licensing materials. Then, in a letter dated September 4, 2024, Meta claimed that because Plaintiffs had six months to clarify the language of RFP No. 45 and did not do so, Meta considers this issue resolved. Not so. The issue will be resolved once Meta conducts a comprehensive search for responsive communications and documents concerning licensing discussions and outreach and produces all relevant material. Please confirm that Meta will search for and produce these materials immediately. Please also confirm that Meta’s search for documents responsive to RFP No. 45—and every other RFP—included all potentially relevant non-custodial data sources in Meta’s possession, custody, or control, and that Meta’s search of custodial data was not limited to just the 10 custodians Meta initially selected.

RFP No. 64:

RFP No. 64 requests “Documents and Communications sufficient to show each instance within the last three years where You have licensed copyrighted works for Meta’s commercial use.” Meta objected “to the phrase ‘commercial use’ as vague, ambiguous, and undefined.” The request is sufficiently clear as to what is being sought. To the extent Meta claims confusion, please explain where the ambiguity lies or why more than one interpretation of the request is reasonable. *Goro v. Flowers Foods, Inc.*, No. 17-CV-02580-JLS-JLB, 2019 WL 6252499, at \*12 (S.D. Cal. Nov. 22, 2019) (holding that Rule 34 “does not require the impossible,” and a “generalized designation should be sufficient when the party seeking discovery cannot give a more particular description and the party from whom discovery is sought will have no difficulty



in understanding what is wanted”). Plaintiffs do not appear to have the full set of executed licenses for training data. This issue remains unresolved and Plaintiffs request that Meta promptly produces all executed licenses for any and all training data.

RFP No. 67:

During the parties’ meet and confer on June 21, 2024, Meta represented that none of its released models could produce output of 25,000 words or longer. Plaintiffs requested that Meta produce documents or communications addressing this claimed limitation on the released models’ capability. To date, Meta has refused to do so on the basis of relevance. Please immediately produce all responsive documents or confirm that the parties are at an impasse so we can raise this issue with Judge Hixson.

**II. Plaintiffs’ Other Identified Issues**

Communications with Business Partners:

Plaintiffs served RFPs for Meta’s communications with Microsoft Corporation, OpenAI, Anthropic, Dell Inc., Cloudflare, Inc., and Qualcomm regarding training data. *See* RFP Nos. 29–34. Plaintiffs have not located any responsive documents in Meta’s productions. These requests are plainly relevant to, among other things, the commercialization of the Llama models, including Meta products incorporating those models, and Plaintiffs expect that these business partners would have discussed training data with Meta. Please confirm Meta will search for and produce all responsive documents immediately or confirm that no such documents exist within Meta’s possession, custody, or control.

Communications with Shadow Libraries:

Plaintiffs served RFPs seeking Meta’s communications with Library Genesis, Z-Library, Sci-Hub, Bibliotik, The Eye, and Anna’s Archive regarding training data. *See* RFP Nos. 7–12. In their August 22, 2024 letter, Plaintiffs again raised the lack of communications with shadow libraries in Meta’s productions. In its September 4, 2024 letter, Meta represented it had not located any communications with shadow libraries based on a “reasonable search.” Please explain what search was conducted, including the search terms used and the non-custodial and custodial data sources searched.

LibGen Training Data (RFP Nos. 1-3 & 7, ROG No. 1):

During the parties’ April 24, 2024 meet and confer, Meta indicated there were no stand alone documents or communications related to Z-Library, Sci-Hub, or similar datasets, asserting that such requests were disproportionate and largely irrelevant under the ESI Order. Plaintiffs narrowed their request to documents related to training data for Llama, and Meta agreed to produce relevant datasets if they were used to train Llama. Please confirm that all such datasets have been produced, including any datasets used for testing, validation and/or benchmarking, and please provide Bates Numbers corresponding to each dataset. Please also confirm that Meta



has produced *all* copies of each dataset that it copied, reproduced and/or used to train any Llama model.

GitHub-Hosted Communications and Source-Code Repositories:

Plaintiffs served RFP Nos. 13, 14, 15, 16, 21, 23, 24, 25, 26, 27, 36, 45, 49, 50, 52, 54, 55, 56, 59, 60, 61, 62, 63, 64, 65, 66, and 67 between December 27, 2023, and April 9, 2024, with objections from Meta served between February 2 and May 29, 2024. Plaintiffs provided Meta with a deficiency letter on these issues on August 22, 2024. During an August 29 meet and confer, Meta indicated that it “fully intended to provide change history, metadata, etc.” for the source code inspection, explaining that while the inspection machine may not offer the same interface as a public GitHub repository, Plaintiffs could use standard git commands to view the change history. Plaintiffs also requested additional metadata including pull requests and issues, and sent another letter to Meta on September 24, 2024 regarding the absence of repository metadata, to which Meta responded that it was unable to identify a specific document request to which Plaintiffs’ requests would be responsive and maintained that the requested data was not relevant. We disagree. *See* RFP Nos. 1–3, 37, 45, and 50; ROG Nos. 1–2, 8, and 13. Meta has yet “to provide [all] change history, metadata, etc.” for its Llama models; the source code and data made available on the workstations are incomplete. Plaintiffs request that Meta make all source code and technical data, including all pull requests and issues, available onsite at Meta for Plaintiffs to inspect, in the manner in which Meta’s employee view such material.

Llama 3, 3.1, 3.2 & Llama 4:

In connection with Plaintiffs’ RFP Nos. 1, 2, 3, and 7, there remains an incomplete production of datasets used to train Llama 3. Meta maintains that datasets beyond what has been produced are not relevant to Plaintiffs’ claims. Not so. During the April 25, 2024 meet and confer, Meta acknowledged it was no longer withholding data for Llama 3 because it had been released, but Meta further stated that it did not plan to collect training data beyond those models. Then, in a May 17, 2024 letter, Meta agreed to produce the actual datasets used to train the released models, noting that the Llama 2 and Llama 3 training datasets overlap. Further, during the August 29, 2024 meet and confer, Meta stated that Plaintiffs should have received all pre-training data for Llama 2 and much of Llama 3, reiterating that it does not believe the Llama 3 datasets are relevant beyond what has been produced, unless Plaintiffs can demonstrate that those datasets contain Plaintiffs’ asserted works. Meta’s September 4 letter confirmed production of the data used to train Llama 1 and Llama 2, including the B3G dataset, which Meta said was also reused for Llama 3, with one additional dataset relating to LibGen, which Meta mailed to Plaintiffs on September 10, 2024.

Despite this back-and-forth over several months, it appears that Meta has not produced all of the requested training data for all of its Llama models—not just Llama 1, Llama 2, and Llama 3, but all versions and including Llama 4—and relating to all aspects of the development of the Llama models, not just the training process but also the pre-training and post-training processes. This data is relevant to the issue of copying and to fair-use factor one because it directly relates to the purpose and character of the use, including whether the Llama models transformed



copyrighted material and whether Meta has used or intends to use used pirated material for a commercial enterprise.

Further, as you know, Meta employee Nikolay Bashlykov downloaded one million English-language books and stored them on Meta's servers, but that material has not yet been produced. Please produce it immediately.

Llama Filters of Training Data (RFP Nos. 45, 64, 74, 76, and 77):

In their August 22, 2024 letter, Plaintiffs identified a tool in Meta's possession, custody, or control that allows for the filtering of training data, which Meta previously indicated did not exist. During the August 29, 2024 meet and confer, and in subsequent correspondence on September 4, 2024, Meta disputed Plaintiffs' characterization. In any event, Plaintiffs request that Meta confirm whether such a tool exists. If any such tool exists, Plaintiffs request that Meta immediately produce it and all associated data and communications.

Valuation and Financials:

Meta claims that there are very few responsive documents in this category but promised to produce a few more; however, it remains unclear whether it has done so. In Plaintiffs' letter dated August 22, 2024, Plaintiffs raised concerns regarding Meta's production, noting a lack of documents related to Meta's actual or projected income from the sale or licensing of the Meta Language Models (RFP No. 46), as well as the absence of financial documents (RFP No. 53). In response to RFP No. 46, Meta claimed these requests seek privileged documents, and are too broad, burdensome, and irrelevant. Additionally, Meta objected to RFP No. 53 on similar grounds, adding that the request wrongly assumed Meta keeps specific financial documents for a language model. Meta stated, "[b]y responding to this Request, Meta does not represent that such documents exist."

During the August 29, 2024 meet and confer, Plaintiffs addressed and reiterated concerns about the lack of income and financial document productions. Meta responded that it had produced at least *one* document containing revenues and revenue projections associated with the generative AI program generally. Meta also mentioned that it would include a *couple* of additional documents related to *direct* revenues received from partnerships with Microsoft Azure and Amazon AWS in an upcoming production, stating that this constituted all Meta agreed to produce on this topic. In response to Plaintiffs' September 4, 2024 letter following up on these documents, Meta referred Plaintiffs to Bates No. Meta\_Kadrey\_00089020 for the projected revenues associated with its generative AI program and stated the Azure data was still being collected.

Please confirm that Meta will produce all documents responsive to RFP Nos. 46 and 53 or identify what objections it intends to stand on so we can address with Judge Hixson Meta's responses and objections to these RFPs.





### Production of Hyperlinks:

The parties agreed on the methodology for production of hyperlinked documents, but it does not appear that Meta is producing all hyperlinked documents. As discussed above, because Meta's production letters do not identify what is contained within associated productions, it is difficult, if not impossible, to identify whether Meta has produced hyperlinked documents and which document(s) they are hyperlinked in. In a letter dated August 22, 2024, Plaintiffs raised concerns that several documents produced by Meta contain hyperlinks to documents that apparently exist on the cloud but are not accessible to Plaintiffs. Six example Bates numbers were provided, and Plaintiffs requested production of those documents.

Then, during a meet and confer on August 29, 2024, Plaintiffs inquired about retrieving the hyperlinked documents. Meta responded that the best course of action would be for Plaintiffs to specify which links they were interested in so that Meta could send those particular documents. Meta further stated that many of the links were not true hyperlinks but rather locations of datasets that had either already been produced separately or were deemed irrelevant by Meta because they had not been used for training. And in Meta's letter dated September 4, 2024, Meta confirmed it was reviewing the hyperlinks identified in Plaintiffs' letter to determine if they pointed to materials that were responsive, not privileged, and had not yet been produced. Meta assured prompt production of any such documents. It is unclear whether any of those documents have been produced, and certainly not all have.

On September 23, 2024, Plaintiff emailed Meta, requesting that Meta produce the hyperlinked documents referenced in the "Background Documentation Section" contained in Meta\_Kadrey\_00055914–15 in preparation for an upcoming deposition. On September 30, 2024, Meta emailed Plaintiffs, indicating that two of the requested hyperlinks had already been produced, although it remains unclear whether these correspond to Plaintiffs' request from September 23:

- The hyperlink in Exhibit 11, discussed at Kambadur Tr. 223:10-22; 224:1-8, has been produced four times under Bates numbers Meta\_Kadrey\_00000250, Meta\_Kadrey\_00048964, Meta\_Kadrey\_00057816, and Meta\_Kadrey\_00063314.
- The hyperlink in Exhibit 29, discussed at Mihaylov Tr. 178:14-22; 179:1-18, was previously produced at Meta\_Kadrey\_00089109, Meta\_Kadrey\_00088770, and Meta\_Kadrey\_00089197, which have since been reproduced with updated redactions, along with additional copies at Meta\_Kadrey\_00093388, Meta\_Kadrey\_00093516, and Meta\_Kadrey\_00093424.

Plaintiffs identified and emailed Meta about yet more hyperlinked documents that it does not appear Meta produced on October 1, October 2, October 3, October 4, October 7, and October 8 2024. Meta has not responded. It is clear that this process is unworkable—as Judge Hixson explained with Meta's imaging problems, it is not Plaintiffs burden but Meta's to ensure its productions are complete, which include producing all hyperlinked documents just like parties have to produce all email attachments, subject to any privilege claim and logging them separately on a privilege log. Please confirm that Meta will review its productions to identify all





non-privileged hyperlinked documents and then produce all such documents (and log any subject to a privilege claim). If your vendor suggests that it cannot identify hyperlinks in documents so Meta can then locate and produce those documents, we are happy to assist your vendor in doing so if you connect us with them.

Timothy Dettmers:

Tim Dettmers is a Meta employee who disclosed Meta's views on the risks of using certain datasets in certain communications with third parties. As you know, the issue of whether Meta can claim privilege over those communications is the subject of Judge Hixson's discovery order (ECF No. 114) and Plaintiffs' pending Rule 72 Objections seeking relief therefrom. Regardless, please let us know if you will agree to add Mr. Dettmers as a document custodian and to allow Plaintiffs to depose him without a court order.

Incorrectly Imaged Documents:

Judge Hixson previously ordered Meta "to reproduce documents that have been incorrectly imaged." ECF No. 196. Please confirm the date by which Meta will comply with this order.

ESI Issues Generally:

Meta's production letters do not identify which of Plaintiffs' RFPs the documents within the productions are responsive to. The failure to do so has made it difficult to assess Meta's responses to and compliance with Plaintiffs' discovery requests, particularly as Meta repeatedly has made additional productions on the eve (and morning) of depositions, as well as other random times without explanation. Please amend your production letters to identify which RFP(s) the documents in the accompanying production(s) pertain to.

Plaintiffs also request that Meta produce its search terms and hit reports, and that the parties exchange all potentially relevant non-custodial and custodial data sources and identify all potentially relevant witnesses. The ESI Order already requires Meta to disclose its search terms and hit reports. *See* ECF No. 101 at 4 ("The Producing Party will then apply the original and new search terms to the target document set and disclose the original and new terms to the Requesting Party, along with a hit report."). Please do so immediately. Further, while we recognize the ESI Order does not require the parties to identify all potentially relevant witnesses and all potentially relevant non-custodial and custodial data sources, we believe doing so now is important to potentially obviate future disputes over the adequacy of searches, custodians, and deponents. If Meta refuses to exchange this information, either informally or with an amendment to the ESI Order, then Plaintiffs will raise this issue with Judge Hixson.

\*\*\*\*

Please let me know a suitable time on Friday, October 11 to discuss these issues.



Plaintiffs reserve all rights, and the exclusion of any issues in this correspondence does not suggest or imply that Plaintiffs have waived any rights thereby, including with respect to issues already in the process of meet and confer discussions, including Privilege Log discussions that are ongoing.

Sincerely,

/s/ Maxwell V. Pritt

Maxwell V. Pritt

Boies Schiller Flexner LLP